

Proper Noun Identification Problem In Georgian Language Based Text Processing

M.Khachidze, M .Tsintsadze, M.Archuadze, G.Besiashvili

e-mails: {[manana.khachidze](mailto:manana.khachidze@tsu.ge), [magda.tsintsadze](mailto:magda.tsintsadze@tsu.ge),
[maia.archuadze](mailto:maia.archuadze@tsu.ge), [gela.besiashvili](mailto:gela.besiashvili@tsu.ge)}@tsu.ge,

Computer Science Department, Faculty of Exact and
natural Sciences , I.Javaxishvili Tbilisi State University.

Chavchavadze Ave. 3 Tbilisi, Georgia

Vast majority of Natural Language Processing tasks contain stage of text initial processing dealing with proper noun identification along with several other procedures (stop word withdraw, stemming, lemmatization). The Identification of Proper nouns is a very important stage of Social Networks produced big data processing. The task stated is relatively easy for such languages that spell their proper names in a certain manner , for instance starting them with a capital letter, but for Georgian language like languages task is more complicated.

In the work provided discuss proper noun removing algorithm for Georgian language based texts. The algorithm represents the variation of "Boolean search", using the database of the proper nouns(names). The algorithm success depends on recall of the database. In addition, a complete cycle of the procedure involves the use of the stemming algorithm developed for Georgian language based texts. The mentioned stemming is original and is customized for Georgian language in accordance of its peculiarities, thus leading to originality and customization of our algorithm for Georgian Language.

References

- [1] Riyad Al-Shalabi, Ghassan Kanaan, Bashar Al-Sarayreh, Khalid Khanfar, Ali Al-Ghonmein, Hamed Talhouni, and Salem Al-Azazmeh. Proper Noun Extracting Algorithm for Arabic Language. International Journal of the Computer, the Internet and Management Vol. 19. No.1 (January-April, 2011) pp 45 -53
- [2] Wesley W. Chu. Data Mining and Knowledge Discovery for Big Data: Methodologies, Challenge and Opportunities. Springer Science & Business Media, Sep 24, 2013 - Computers - 311 pages
- [3] Ali Reza Ebadat, Vincent Claveau, Pascale Sebillot. Proper Noun Semantic Clustering Using Bag-of-Vectors. Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference. Marco. May 23, 2012 – May 25, 2012. Pp. 220-225.